
L'archiviste et la gestion des jeux de données ouvertes de recherche en milieu universitaire canadien : une visibilité à accroître, des collaborations à développer

The Archivist and the Management of Open Research Datasets in Canadian Universities: Increasing Visibility and Developing Collaborations

Siham Alaoui

Gestion de l'information documentaire, Département d'information et de communication, Université Laval, Québec, Canada
(siham.alaoui@com.ulaval.ca)

Abstract

In an academic world advocating for open science values, researchers should deposit their research datasets in a form that is open and reusable by users. Provided that these data are reusable, it is important that they meet some criteria defined according to the FAIR (*Findable, accessible, interoperable, and reusable*) principles. To this end, a number of experts, namely academic librarians, help researchers manage their research datasets. From an archival point of view, the latter are viewed as evidence of scientific activities. Despite having necessary expertise to manage these data, archivists are not always present in this context and if they are, their role is limited to the long-term preservation of these data. In this article, open research data management is studied from an archival perspective, conceptualizing the role of university archivists in the governance of these traces. By mobilizing relevant scientific articles in information science, we model the relationships archivists should develop with other stakeholders involved in open research data governance in Canadian universities. Our research contributes to advancing knowledge about the role of university archivists in collaborative research data governance.

Keywords: open research data governance; archivists; librarians; open science; scientific transparency; universities; Canada

Résumé

Avec le mouvement de la science ouverte militant pour un accès équitable aux résultats de la recherche scientifique, les chercheurs universitaires sont tenus de déposer leurs jeux de données de recherche sous une forme ouverte et réutilisable par un tiers. Pourvu que ces traces documentaires soient réexploitables, il importe qu'elles obéissent à des critères de qualité définies selon les principes FAIR (*Findable, accessible, interoperable and reusable*). Pour ce faire, plusieurs experts, notamment les bibliothécaires universitaires, accompagnent les chercheurs dans la gestion de leurs jeux de données de recherche. D'un point de vue archivistique, les jeux de données de recherche sont des preuves des activités scientifiques, ce qui laisse entendre qu'elles devraient également être prises en charge par les archivistes universitaires. Cependant, ces derniers sont souvent peu présents dans ce contexte et leur rôle est, dans la plupart des cas, limité à la préservation à long terme de ces données. Dans cet article, nous nous attachons à la gestion des jeux de données ouvertes de recherche d'une perspective archivistique, en conceptualisant le rôle des archivistes universitaires dans la gestion collaborative de ces traces. En mobilisant des corpus scientifiques pertinents en sciences de l'information, nous modélisons les relations que les archivistes entretiennent avec les autres parties prenantes impliquées dans la gouvernance des données ouvertes de recherche en milieu universitaire canadien. Notre recherche contribue à l'avancement des connaissances sur le rôle de l'archiviste universitaire dans la gouvernance collaborative des données de recherche.

Mots clés : gouvernance des jeux de données ouvertes de recherche; archivistes; bibliothécaires; science ouverte; transparence scientifique; universités; Canada.

Introduction

Avec le développement du numérique et des pratiques axées sur la gouvernance ouverte, les organisations publiques sont appelées à faire preuve d'une ouverture de leur information documentaire envers les communautés qu'elles desservent. L'établissement du *Partenariat pour un gouvernement ouvert*¹ en 2011 a comme vocation de promouvoir les valeurs de transparence dans les organisations pour favoriser une prise de décision collective, fondée sur des données probantes. Depuis 2012, le Canada est membre de ce Partenariat² et s'est engagé pour promouvoir ces valeurs pour bâtir une société ouverte, inclusive et transparente, et ce, en définissant un ensemble d'objectifs à atteindre dans une période déterminée. Les engagements du Canada concernent non seulement les activités administratives, mais aussi scientifiques. Soulignons que la question de l'accès équitable au savoir est présente depuis des années dans le monde de la recherche et que les engagements du Canada s'inscrivent dans une optique du renforcement de ces valeurs dans les universités. En fait, dans le *Plan d'action du Canada pour un gouvernement ouvert (2014-2016)*³, le pays a défini, parmi ses objectifs, celui de « maximiser l'accès à la recherche scientifique financée par le gouvernement fédéral afin de favoriser une collaboration et un dialogue accrus avec la collectivité scientifique, le secteur privé et le public » (Gouvernement du Canada, s.d.). Quelques années plus tard, le *Plan d'action pour la science ouverte 2021-2026*⁴ a été publié pour répondre à ces préoccupations de l'accès équitable aux résultats de la recherche scientifique pour ensuite les mobiliser dans le but de trouver des solutions aux problèmes de la société canadienne, tout en favorisant une synergie et une collaboration avec les milieux universitaires (Gouvernement du Canada, s.d.).

Ainsi, en tant qu'établissements publics ayant pour missions l'enseignement et la recherche, les universités canadiennes sont concernées par ces ambitions de la transparence scientifique. Les chercheurs universitaires sont soumis à des impératifs de transparence, notamment pour leurs projets de recherche financés par les organismes subventionnaires de la recherche du Canada, dont les Instituts de recherche en santé du Canada (IRSC), le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) et le Conseil de recherches en sciences humaines (CRSH), en prouvant à ceux-ci la manière dont les fonds alloués ont été exploités dans le cadre de leurs activités de recherche. C'est ainsi que

¹ Voir : <https://www.opengovpartnership.org/fr/> (page consultée le 12 juin 2025)

² Voir : <https://www.canada.ca/fr/gouvernement/systeme/rapports-pangouvernementaux-depenses-activites/confiance-transparence/a-propos-gouvernement-ouvert/plan-action-national-gouvernement-ouvert.html> (page consultée le 10 juin 2025)

³ Voir : <https://ouvert.canada.ca/fr/contenu/plan-daction-du-canada-gouvernement-ouvert-2014-2016> (page consultée le 12 juin 2025). Le gouvernement ouvert se base sur trois principes, à savoir la transparence administrative et scientifique, la participation et la collaboration citoyennes. Il vise à renforcer la relation des citoyens avec l'État grâce à leur implication dans la prise de décision publique.

⁴ Voir : <https://www.canada.ca/fr/environnement-changement-climatique/services/sciences-technologies/plan-action-science-ouverte.html> (page consultée le 12 juin 2025)

la *Politique des trois organismes subventionnaires sur le libre accès aux publications*⁵ a été publiée en 2015 pour baliser ces devoirs des chercheurs dans la transparence des activités de recherche.

Par ailleurs, la crise de la COVID-19 a multiplié ces pressions de transparence scientifique en vue de répondre aux enjeux sanitaires et sociétaux en mobilisant des résultats de la recherche scientifique (Unesco, s.d.). Les organismes subventionnaires fédéraux se sont aussi dotés d'une politique encadrant la libération des jeux de données ouvertes de recherche. Il est question des jeux de données brutes, collectées dans le cadre des projets de recherche par le biais des observations, des sondages, des entrevues et des expérimentations cliniques, entre autres, et qui peuvent être réexploités par une tierce partie selon des conditions spécifiées dans une licence ouverte (ex. *Creative Commons*) (Gouvernement du Canada, s.d.). La politique, intitulée la *Politique des trois organismes subventionnaires sur la gestion des données de recherche*⁶, stipule que les chercheurs titulaires de projets financés par les trois organismes doivent rendre publiquement accessibles leurs jeux de données de recherche sous une forme ouverte sur un dépôt institutionnel dédié à cette fin (Gouvernement du Canada, s.d.). Plus spécifiquement, le dépôt *Borealis*⁷ est le dispositif technologique utilisé par les universités canadiennes pour assurer une meilleure diffusion de ces jeux de données par les chercheurs, les étudiants à la maîtrise et au doctorat, les stagiaires postdoctoraux et les professionnels de recherche.

Considérant les compétences variables de ces acteurs concernés par la libération des données de recherche, ils sont accompagnés par un ensemble de personnes spécialistes qui leur offrent un service-conseil pour que les données libérées soient réexploitables pour soutenir les ambitions de la recherche scientifique. Il s'agit des bibliothécaires universitaires⁸, des analystes des données et des experts en technologies de l'information (TI) qui accompagnent les producteurs, et ce, tout au long du cycle de vie de ces données (Da Sylva et al., 2019). À ces parties prenantes s'ajoutent les conseillers en recherche et en conformité qui s'attachent à des aspects particuliers entourant la réutilisation des données en fonction des contraintes éthiques et des ententes avec les partenaires de la recherche (Alaoui, 2024; Barsky et al., 2023; Da Sylva et al., 2019). Dans certains cas, les archivistes

⁵ Voir : <https://cihr-irsc.gc.ca/f/32005.html> (page consultée le 12 juin 2025)

⁶ Voir : <https://science.gc.ca/site/science/fr/financementinterorganismes-recherche/politiques-lignes-directrices/gestion-donneesrecherche/politique-trois-organismes-gestion-donnees-recherche> (page consultée le 12 juin 2025)

⁷ <https://borealisdata.ca/> (page consultée le 12 juin 2025)

⁸ Les bibliothécaires universitaires sont, comme leur titre l'indique, des professionnels de l'information rattachés à des bibliothèques universitaires. Leur rôle est de faciliter l'accès à l'information scientifique et technique pour les membres de la communauté universitaire. Ils assurent en outre le développement des compétences informationnelles du corps enseignant et étudiant pour repérer et exploiter l'information dont ils ont besoin.

universitaires⁹ sont aussi interpellés dans la gestion de ces jeux de données de recherche, notamment pour ce qui a trait à la préservation à long terme de ces traces documentaires et leur évaluation¹⁰ (Alaoui, 2024; Borglund et Borgerud, 2021; Makhoulf-Shabou et Echernier, 2017). Cependant, ces rôles s'avèrent limités, puisque ces experts jouissent d'un ensemble de compétences qui leur permettent d'accompagner les producteurs dans la gestion de ces données, et ce, depuis leur création jusqu'à leur élimination ou leur préservation à long terme, en couvrant ainsi d'autres interventions archivistiques allant au-delà de la préservation à long terme.

Dans cette contribution, nous désirons aborder la question de la gouvernance des jeux de données ouvertes de recherche d'un point de vue archivistique. Nous nous attachons aussi aux avenues de collaboration que les archivistes universitaires pourraient développer avec les autres acteurs de l'écosystème de la gestion des données ouvertes de recherche, notamment pour encourager les valeurs inhérentes au mouvement de la science ouverte. Pour ce faire, nous commençons par une revue de littérature qui expose la cartographie des acteurs impliqués dans un tel contexte, avant de nous attacher à la perspective archivistique de la gestion de ces traces documentaires. S'ensuivra une analyse critique de certains corpus scientifiques en vue d'apporter de nouveaux ancrages théoriques au rôle des archivistes universitaires dans la gestion des jeux de données ouvertes de recherche.

Méthodologie

Afin de répondre à notre objectif, nous allons mobiliser un ensemble d'écrits scientifiques pertinents en archivistique, notamment ceux qui abordent la notion du document d'archives (*record*) dans le contexte numérique. L'idée est de justifier la pertinence de l'intervention archivistique dans la gestion des jeux de données ouvertes de recherche comme traces authentiques, fiables, intègres et exploitables, rendant compte des activités scientifiques menées à l'université. S'ensuivra une lecture fonctionnelle des jeux de données de recherche d'un point de vue archivistique, en mettant de l'avant les finalités de leur exploitation sous l'angle institutionnel, opérationnel et symbolique. Un tel exposé nous amènera à conceptualiser les rôles que les archivistes universitaires pourraient jouer dans la gestion de ces traces, notamment sur le plan de leur contextualisation et évaluation. Pour ce faire, une nouvelle lecture des valeurs archivistiques que ces jeux de données pourraient avoir sera établie sous l'angle fonctionnel. Une telle démonstration permettra d'illustrer les valeurs plurielles de ces traces, d'où le nécessaire dialogue entre les archivistes, les producteurs, les usagers et les experts pour déterminer, sur une base collaborative, les jeux de données à conserver en fonction de leurs finalités d'usage à l'université. Notons que

⁹ Les archivistes universitaires ont pour rôle d'assurer la traçabilité des activités de l'université et la constitution de sa mémoire organique et consignée témoignant de l'évolution de ses missions au fil du temps.

¹⁰ L'évaluation des archives consiste à déterminer la valeur de ces traces documentaires sur le plan administratif, légal et symbolique (Schellenberg, 1956). Elle vise à optimiser la gestion des documents pour n'en préserver que ceux qui sont utiles pour l'organisation et ses publics externes (ex. chercheurs, journalistes, grand public, partenaires, etc.).

notre démonstration se fonde sur des constats propres à la réalité de la gestion des données de recherche dans le contexte universitaire canadien.

La libération des données ouvertes de recherche : un aperçu à la lumière des écrits existants

Le concept des données ouvertes (*open data*) n'est pas nouveau. Les chercheurs universitaires ont depuis longtemps milité pour une diffusion des données statistiques sous forme de bases de données publiques destinées à être réutilisées par les communautés scientifiques (Kitchin, 2014). Il est question des données ouvertes de recherche, lesquelles renvoient aux données issues des activités de recherche scientifique (Castets Renard et Gandon, 2016; Reymonet, 2017). Selon l'Organisation de la Coopération et du Développement Économique (OCDE), les données ouvertes de recherche sont définies comme des enregistrements factuels (ex. chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour la recherche scientifique (OCDE, 2007). Il peut s'agir des carnets de laboratoire, des notes de terrain, des analyses préliminaires, des rapports de recherche, etc. Le concept des données de recherche s'est popularisé dans les années 2000 avec l'engouement pour le mouvement du libre accès (*open access*), des archives ouvertes (*open archive*) et de la science ouverte (*open science*) (Jacquemin et al., 2019; Quarati et Raffaghelli, 2020).

La libération des jeux de données de recherche sous une forme ouverte répond aux pressions en lien avec la transparence scientifique et l'imputabilité, notamment avec le développement du discours sur le libre accès aux résultats de la recherche (Awada et al., 2021). Elle soutient en outre la visibilité des recherches, stimule la collaboration entre les chercheurs et promeut la réutilisation équitable des résultats scientifiques à des fins de création de la valeur publique, scientifique et économique (Da Sylva et al., 2019; Dillaerts, 2020; Quarati et Raffaghelli, 2020). La diffusion des résultats d'enquêtes, d'expérimentations et d'essais de laboratoires sur le portail canadien *DataVerse* constitue un exemple du désir de la communauté scientifique canadienne de soutenir le rayonnement de la recherche universitaire et à l'avancement des connaissances dans différentes disciplines scientifiques. Pour ce faire, les données de recherche doivent répondre à un ensemble d'exigences spécifiées par Wilkinson et ses collaborateurs (2016), connus sous l'acronyme FAIR (*Findable, Accessible, Interoperable et Reusable*) et qui correspondent, respectivement, à la découvrabilité des données à l'aide des métadonnées; l'accessibilité de ces traces sur des dépôts publics; leur interopérabilité grâce à l'usage des formats technologiques ouverts; et la réutilisabilité¹¹ des données de recherche suivant des conditions définies dans des licences d'utilisation (Wilkinson et al., 2016). En vue de souscrire à ces exigences, les producteurs des données de recherche sont tenus d'élaborer des plans de gestion des données (*Data management plan*). Un tel outil de gestion fournit

¹¹ La réutilisabilité renvoie à l'appropriation des données dans des contextes autres que ceux de leur création initiale. Ces données peuvent être agrégées à d'autres objets documentaires pour créer de nouvelles traces à plusieurs fins.

des renseignements sur le contexte de la recherche ayant donné naissance aux données en question, la portée de celles-ci et les modalités de leur gestion, ainsi que les conditions de leur réutilisation (Gouvernement du Canada, 2021).

En bref, la diffusion des jeux de données de recherche sous une forme ouverte sur des portails publics tels que *Dataverse* s'inscrit dans la continuité des efforts déployés pour promouvoir un meilleur accès des résultats de la recherche, ce qui aide les chercheurs à innover et développer des opportunités de recherche en évitant de réinventer la roue. Pour ce faire, la gestion de ces jeux de données doit s'opérer dans le respect des principes FAIR, ce qui met en jeu plusieurs parties prenantes qui possèdent l'expertise nécessaire à cette fin.

Les acteurs de la gestion des données ouvertes de recherche

L'atteinte des ambitions de la science ouverte en favorisant la réutilisation des jeux de données de recherche exige que leur création, leur gestion, la diffusion et la préservation à long terme fassent l'objet d'une gouvernance appropriée. Calquée sur l'univers des données, la gouvernance des données renvoie aux pratiques et aux mécanismes par lesquels la création, le traitement, la diffusion et la préservation pérenne des données sont assurés conformément à un cadre de bonnes pratiques, de normes, de politiques et de règlements (Sargiotis, 2024). La gouvernance des données ouvertes de recherche met en jeu les professeurs-chercheurs universitaires, les étudiants aux cycles supérieurs, les professionnels de recherche et les stagiaires postdoctoraux (Alaoui, 2024; Da Sylva, 2021; Da Sylva et al., 2019) comme producteurs de ces traces. Les professeurs-chercheurs dirigent par eux-mêmes leurs activités de recherche et celles des étudiants aux cycles supérieurs, des professionnels de recherche et des stagiaires postdoctoraux sous leur supervision. Ils procèdent à la mise en ligne des jeux de données ouvertes qu'ils produisent dans le cadre de leurs projets (Da Sylva, 2021; Da Sylva et al., 2019; Makhoul-Shabou et Echernier, 2017). Les étudiants, les stagiaires postdoctoraux, de même que les professionnels de recherche, sont également des producteurs des données ouvertes de recherche, puisque celles-ci découlent de leurs travaux supervisés par les professeurs-chercheurs. Les organismes subventionnaires de la recherche élaborent les politiques qui encadrent la gestion des données ouvertes de recherche découlant des projets qu'ils subventionnent, comme c'est le cas pour leur politique établie en 2021. Dans cette même optique de gouvernance des données ouvertes de recherche, les comités éthiques de la recherche, souvent exerçant dans trois secteurs (multidisciplinaire, sciences de la santé et psychologie), déterminent les mécanismes de gouvernance sur le plan éthique, notamment pour ce qui a trait aux questions du droit d'auteur et de la protection de la vie privée. À ce sujet, ils veillent à ce que les formulaires de consentement accompagnant la démarche de collecte des données auprès des participants permettent l'autorisation de la diffusion et la réutilisation des données de recherche à des fins secondaires (Alaoui, 2024). Ils évaluent aussi le risque de croisement des données et se prononcent sur les mesures sécuritaires pour assurer la protection de la vie privée des participants. Ils peuvent aussi influencer les durées de conservation des données ouvertes de recherche, conformément aux règles éthiques et légales en vigueur.

Dans un objectif de mieux aider les producteurs des données de recherche dans leur démarche d'opérationnalisation des directives énoncées dans la *Politique des trois organismes de recherche*, plusieurs spécialistes interviennent dans cette optique. Les bibliothécaires universitaires ont, tout naturellement, un rôle conseil dans ce contexte, puisque les données de recherche découlent des publications mêmes des producteurs (Da Sylva, 2017). Plus précisément, le rôle de ces professionnels s'inscrit dans un large contexte de démocratisation d'accès au savoir et d'établissement des balises de l'accès équitable à l'information scientifique, notamment sous l'angle du respect de la propriété intellectuelle de l'information documentaire à rendre publiquement accessible (Lapointe et Castonguay, 2020). Les bibliothécaires sont interpellés à titre de conseillers dans la gestion des jeux de données ouvertes de recherche (Da Sylva, 2017; Makhoul-Shabou et Echernier, 2017), puisque ces derniers sont décrits en fonction de leur portée thématique. Cela laisse entendre que ces traces sont traitées comme des publications plutôt que des produits de l'activité. Qui plus est, l'intervention des bibliothécaires auprès des producteurs des données se fait en amont, notamment en mettant à la disposition de ces derniers des modèles de plans de gestion des données à utiliser pour expliciter les modalités de traitement, d'exploitation et de préservation de ces données. Ils rédigent aussi des manuels et des guides, de même que des conventions de nommage et des schémas de métadonnées, afin de décrire les jeux de données de recherche en fonction de leur portée thématique (Alaoui, 2024). En clair, l'intervention des bibliothécaires vise à offrir aux producteurs des données l'accompagnement nécessaire pour que ces traces documentaires répondent aux principes FAIR.

D'autres spécialistes interviennent dans la gouvernance des données ouvertes de recherche. Il est question des experts en TI qui ont un rôle de conseil dans la conception et le déploiement des infrastructures technologiques soutenant la création, la curation, la diffusion et la préservation à long terme des données de recherche. Ils déploient en outre les dépôts de données de recherche et s'assurent de leur utilisabilité et ergonomie, Ils offrent aussi des recommandations en regard de l'interopérabilité des formats de libération et de préservation à long terme de ces données pour contrer l'obsolescence technologique et les risques d'instabilité du contenu (Alaoui, 2024; Da Sylva, 2021). Dans cette même logique de soutien, des conseillers dans des aspects particuliers en lien avec la gouvernance des données ouvertes de recherche interviennent dans cette optique. Il est question des conseillers en conformité, en développement et en valorisation de la recherche (Alaoui, 2024). Les conseillers en conformité, notamment en droit d'auteur, s'attachent aux questions en lien avec la propriété intellectuelle et la définition des modalités du respect du droit d'auteur, lesquelles sont balisées par des licences d'utilisation du type *Creative Commons*. Dans cette même lignée de pensée, les conseillers en développement et en valorisation de la recherche s'attachent aux conditions de libération des données de recherche, notamment en négociant les clauses avec les partenaires des chercheurs universitaires, pour aboutir à des conditions de diffusion et de réutilisation qui respectent les clauses des ententes de partenariats dans le cadre du financement des projets de recherche notamment par des partenaires et des firmes de l'industrie (Alaoui, 2024).

Le rôle des archivistes universitaires dans la gestion des données de recherche est souligné dans certains écrits, notamment en lien la préservation à long terme de celles-ci (Da Sylva, 2021; Makhoulf-Shabou et Echernier, 2017; Quarati et Raffeghelli, 2020). Or, peu d'études abordent le rôle de ces experts dans la contextualisation de ces traces, bien que les jeux de données de recherche doivent être interprétés dans leur contexte de production scientifique (Da Sylva, 2017). En fait, le concept de données ouvertes de recherche est abordé, dans la plupart des écrits recensés, dans son acception bibliothéconomique plutôt qu'archivistique. Les chercheurs n'abordent pas la libération des jeux de données ouvertes comme des documents (*records*) au sens archivistique du terme, mais plutôt comme des publications, en oubliant qu'un jeu de données est aussi un document au sens archivistique (Borglund et Borgerud, 2021). Il importe dès lors d'aborder les jeux de données ouvertes de recherche d'une perspective archivistique afin de justifier l'importance de l'intervention des archivistes universitaires dans la gouvernance de ces traces.

Les données ouvertes de recherche : la perspective archivistique

La perspective archivistique considère le document comme étant un produit d'une activité (Cardin, 2013-2014; Duranti et Thibodeau, 2006). Non seulement rend-il compte de cette activité, mais en permet la réalisation effective dans une situation de transaction donnée. Ainsi, le document d'archives est un produit de l'activité (*by product*) qui se recycle continuellement dans plusieurs contextes pour accomplir plusieurs tâches (InterPARES, s.d.). Dans l'univers numérique, il est aussi agrégé à d'autres objets documentaires dans différents contextes pour atteindre plusieurs finalités d'usage. Dans le contexte de la recherche scientifique, les chercheurs produisent et reçoivent plusieurs types de documents soutenant leurs activités de recherche. Il peut être question des notes personnelles, des études préliminaires et des fiches de lecture, comme il peut s'agir des articles, des rapports de recherche, des thèses et des mémoires, de même que les actes de congrès et les communications scientifiques. Notons que bien que ces documents soient habituellement pris en charge par les bibliothécaires en tant que publications, ils ont un statut archivistique car ils servent à la fois de preuve des activités des recherches, de véhicules de l'expertise en matière de conduite des projets de recherche en recherche, et de moyens de transfert des valeurs de liberté académique.

Qu'en est-il des données qui découlent de ces documents ? Comment pourrait-on les qualifier d'archives (*records*)? D'un point de vue archivistique, un document d'archives résulte de l'inscription d'une information sur un support plus ou moins stable. Rappelons qu'une information est un ensemble de données traitées pour en extraire du sens. Ainsi, les données en elles-mêmes ne sont pas des documents, mais les jeux de données (*datasets*) peuvent être considérés comme tels, puisqu'ils résultent d'une combinaison logique d'un ensemble de données mises en lien les unes avec les autres pour former un tout cohérent assimilable à un document avec un contenu stable. Ainsi, pour être pris en charge par les archivistes comme preuves des activités, il importe de considérer les données comme des documents, autrement dit, ce sont les jeux de données qui constituent des objets de gestion de l'archiviste. Ces documents d'archives sont soit des documents de gestion ou d'exploitation (de fonction). Les documents de gestion sont ceux qui sont communs à toutes les organisations, puisqu'ils sont produits ou reçus dans le cadre des activités communes telles que la gestion des ressources humaines, financières et informationnelles (Couture et Roy, 2006-2007). Les documents de fonction ou encore d'exploitation sont

ceux qui découlent des missions propres de l'organisation (Couture et Roy, 2006-2007), ici l'université, soit l'enseignement et la recherche. Puisque les jeux de données ouvertes de recherche découlent de l'une des missions propres des universités, à savoir la recherche, ils sont considérés comme des documents d'exploitation, représentant ainsi les résultats des recherches sous forme de données primaires, telles que celles des sondages, les transcriptions des entrevues et les notes sur le terrain.

Pourquoi est-il important de considérer les jeux de données ouvertes de recherche d'une perspective archivistique ? En fait ces traces documentaires servent, en tant que documents d'archives, à atteindre plusieurs finalités inhérentes au triple pouvoir des archives, à savoir la transparence et la défense des droits, l'amélioration de la performance organisationnelle et la reproduction culturelle (Cardin, 2013-2014). Sur le plan de la transparence et de la défense des droits, la mise en ligne des jeux de données ouvertes de recherche constitue une preuve de légitimité des chercheurs, des étudiants, des stagiaires postdoctoraux et des professionnels de recherche en regard des exigences institutionnelles liées à la transparence scientifique et à la science ouverte. D'un point de vue institutionnel, ces acteurs appartiennent à une structure de régulation, puisque l'université est vue comme une institution qui se dote d'un ensemble de règles, de politiques et de mécanismes de gouvernance encadrant ses activités selon des principes communément partagés (Toulouse, 2007). Dans le cadre institutionnel, ces acteurs créent, gèrent et diffusent ces jeux de données de recherche sous une forme ouverte pour prouver aux organismes subventionnaires la manière dont les ressources allouées sont exploitées dans le cadre des projets de recherche subventionnés (Alaoui, 2024). Ainsi, la diffusion des jeux de données de recherche sous une forme ouverte soutient la gestion de l'image de l'université comme institution qui encourage les valeurs de transparence scientifique et d'accès équitable au savoir produit par les membres de sa communauté.

Par ailleurs, les jeux de données peuvent être perçus comme un moyen de transmission de l'expertise de recherche scientifique. En fait, d'un point de vue opérationnel, l'université peut être considérée comme une organisation (Mintzberg, 1985) misant sur l'amélioration de sa performance sur le plan scientifique et éducatif. Cela illustre une quête de rendement économique, de génération de la valeur ajoutée en vue d'attirer des étudiants et des collaborateurs, tout en soignant l'image de marque de l'université en tant qu'établissement d'enseignement supérieur de haut calibre. La libération des jeux de données ouvertes de recherche par leurs producteurs soutient la visibilité de ces derniers sur le plan scientifique. Plus précisément, les portails dédiés à la diffusion de ces traces documentaires ouvertes servent de vitrine reflétant l'expertise universitaire sur le plan scientifique, ce qui favorise la constitution d'une image de marque de qualité des universités. Qui plus est, la mise en ligne de ces jeux de données s'inscrit dans une perspective d'amélioration de la performance des producteurs de ces traces sur le plan scientifique, puisque cela leur permet d'accéder à des données de recherche de qualité qui pourraient par la suite être remobilisées dans d'autres recherches scientifiques. Cela favorise une réutilisation plus optimale des ressources, du temps et de l'énergie, ce qui s'inscrit dans une logique rationnelle d'utilisation des moyens mis en place. En bref, la diffusion des jeux de données ouvertes de recherche aide les producteurs à améliorer leur performance universitaire sur le plan scientifique, notamment en misant sur une utilisation rationnelle du temps et des ressources.

Enfin, la mise en ligne des jeux de données ouvertes de recherche sur des portails publics soutient des ambitions communautaires. En fait, sur le plan socioculturel, l'université est

perçue comme une communauté de pairs (Toulouse, 2007), réunis par des principes de collégialité et de liberté académique. Dans le monde scientifique, les professeurs, les étudiants, les professionnels de recherche et les stagiaires postdoctoraux, forment des regroupements défendant des intérêts communs et poursuivent des finalités mutuelles sur le plan scientifique. Les groupes de recherche, les communautés scientifiques et les comités par disciplines scientifiques en sont quelques exemples. Les documents que ces regroupements produisent ou reçoivent dans le cadre de leurs activités de recherche, et qui découlent directement de celles-ci, sont des moyens de transmission des valeurs de liberté académique, de critères de rigueur scientifique reconnus par les membres de ces communautés. Ces valeurs et principes favorisent la collaboration entre les regroupements pour faire avancer la recherche, requestionner certaines pratiques, et développer des avenues de collaboration dans une perspective de recherche interdisciplinaire et multidisciplinaire. Dans cet esprit, l'ouverture de ces jeux de données de recherche soutient l'implication auprès des communautés scientifiques, la reconnaissance des exploits sur le plan scientifique, le développement des relations de collaboration entre les chercheurs ainsi que la transmission des valeurs scientifiques. La disponibilité en ligne de ces traces documentaires soutient des valeurs telles que la confiance et l'intégrité scientifique, lesquelles fondent les échanges et les relations entre les pairs, unis par des principes de collégialité.

En somme, les jeux de données de recherche sont des documents d'archives, puisqu'ils sont des produits des activités de recherche qui se recyclent continuellement dans plusieurs contextes scientifiques. Ils soutiennent des finalités à la fois institutionnelles, organisationnelles et communautaires. Celles-ci renvoient au triple pouvoir des archives (défense des droits, amélioration de la performance et valorisation des liens communautaires (Cardin, 2013-2014; Senécal, 2007-2008; Shepherd et Yeo, 2013). L'insertion des jeux de données de recherche dans le discours archivistique permet non seulement de définir ces traces comme des documents d'archives (*records*), mais aussi de mettre de l'avant l'importance de l'intervention archivistique dans le cadre de leur gestion. En fait, celle-ci, malgré ses bénéfices, présente des défis archivistiques qu'il importe de pallier par des interventions appropriées.

Vers de nouveaux ancrages théoriques sur le rôle de l'archiviste universitaire dans la gouvernance des jeux de données ouvertes de recherche

La gestion des données de recherche se fait en fonction d'un cycle de vie s'opère d'une manière circulaire, étant donné la nature même de ces traces qui sont facilement agrégeables à d'autres données et réutilisables dans différents contextes. Ce cycle de vie comporte plusieurs étapes correspondant aux interventions faites sur ces données, à savoir la création, le traitement, l'analyse, l'archivage, la diffusion et la réutilisation (voir figure 1).

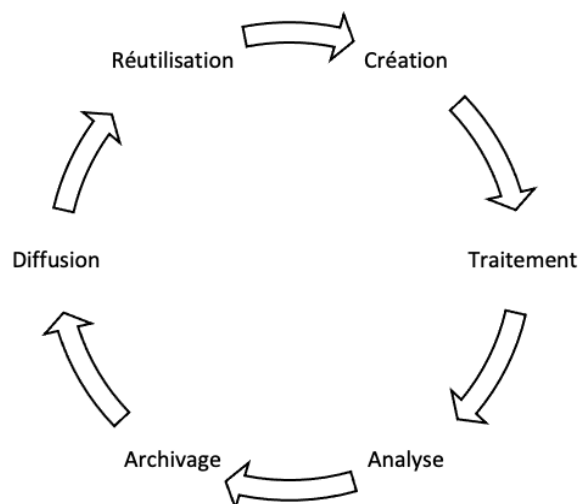


Figure 1 Le cycle de vie des données de recherche (Inspiré de Higgins, 2012)

Les données de recherche sont créées par les professeurs, les étudiants, les professionnels de recherche et les stagiaires postdoctoraux dans le cadre de leurs activités scientifiques. Dès que ces acteurs obtiennent des données collectées via diverses méthodes, dont les sondages, les entrevues, la sélection de la documentation pertinente, l'observation sur le terrain, ou encore l'expérimentation dans des laboratoires, ces traces brutes sont combinées pour former des jeux de données intelligibles. Ces derniers sont ensuite catégorisés en fonction de leur typologie (qualitative vs quantitative). Considérant les problèmes auxquels ces acteurs peuvent se confronter, notamment en termes de manque de données pour certaines questions de recherche, ou encore le biais ou l'incohérence introduits dans ces données, il importe que celles-ci subissent un processus de traitement préalable avant d'être soumises à l'analyse. À l'étape de traitement des jeux de données, ils sont nettoyés, décrits et curés par les producteurs, pour qu'elles soient déposées dans un entrepôt de données disciplinaire ou multidisciplinaire (ex. le *Dépôt fédéré de données de recherche*¹², *Zenodo*¹³). Les doublons étant éliminés et les incohérences soulevées et corrigées, les données sont maintenant prêtes à être analysées par les chercheurs en vue d'atteindre les ambitions de leurs projets de recherche. Elles subissent, tout dépendamment de leur typologie, un processus d'analyse qualitative ou quantitative, tout en prenant en considération les principes de rigueur scientifique régissant chaque discipline. Conscients de l'importance de la traçabilité des activités de recherche et soucieux de la conformité aux exigences de la *Politique des trois organismes subventionnaires de la recherche du Canada*, les chercheurs, titulaires de subventions de la part des organismes finançant la recherche, définissent les modalités d'archivage à court, à moyen et à long terme de leurs données de recherche, ainsi que les délais de conservation de celles-ci. Elles sont ensuite rendues disponibles sur des portails publics, tels que *Borealis* pour les universités canadiennes, pour être réutilisées par un tiers à des fins de recherche, d'enseignement, de vérification de conformité, ou encore d'information générale.

¹² Voir : <https://www.frdr-dfdr.ca/repo/> (page consultée le 12 juin 2025)

¹³ Voir : <https://zenodo.org/>

Considérant la disparité au niveau des compétences des chercheurs dans la curation de leurs données de recherche (Alaoui, 2024; Da Sylva, 2021; Makhoul-Shabou et Echernier, 2017), plusieurs experts accompagnent ces derniers dans la gouvernance des données ouvertes de recherche. Tel que précédemment évoqué, il est question des bibliothécaires, des experts en TI et des analystes des données qui accompagnent les producteurs pour assurer la gestion de leurs données conformément aux principes FAIR (Da Sylva, 2021) selon des conditions spécifiées dans des plans de gestion de données¹⁴. Ces outils, constituant un résultat d'une fédération de l'expertise de plusieurs spécialistes, définissent les modalités de gestion, de traitement, de diffusion et de préservation à long terme de ces données. Or, l'approche de description des jeux de données, par exemple ceux libérés sur les portails publics tels que *Borealis*, est essentiellement de nature bibliothéconomique, puisque ceux-ci sont souvent décrits en tant que publications (à l'aide des métadonnées comme l'auteur, le titre du projet, le titre de la subvention, le nom de l'organisme subventionnaire, des mots-clés thématiques, la licence d'utilisation, la date de diffusion et de mise à jour), les coordonnées des chercheurs et des partenaires du projet en question, c'est-à-dire à en tant que documents autonomes (Voir figure 2). Dans certains cas, ces jeux de données sont mis en lien les uns avec les autres, mais ces relations se limitent souvent à des aspects bibliographiques, par exemple la mention de la publication dont les jeux de données découlent, les articles citant les ces derniers, ainsi que les identifiants ORCID (*Open Research and Contributor ID*) des auteurs, menant à des travaux – essentiellement des articles et des rapports de recherche – qui sont basés sur ces jeux de données.

¹⁴ Voir l'exemple suivant pour explorer les composantes qu'on devrait habituellement renseigner dans un plan de gestion des données : <https://zenodo.org/records/4495482> (page consultée le 12 juin 2025)



Figure 2 Liste des métadonnées à associer aux jeux de données de recherche en vue de leur publication sur *Borealis* (Borealis Université Laval, s.d.)

Toutes ces considérations nous amènent à affirmer que la description des jeux de données ouvertes de recherche ne tient pas toujours compte des activités ayant présidé à leur création, mais plutôt de leur dimension thématique. Autrement dit, les jeux de données ne sont pas décrits comme produits de l'activité de recherche, une description qui sollicite l'expertise des archivistes universitaires. Pourvu que ceux-ci soient plus activement impliqués à cette phase du cycle de vie des données de recherche, il importe de définir les modalités de la gestion de ces traces selon les procédés archivistiques. Cela implique leur description en dépassant le contenu et le contenant pour couvrir également le contexte, ainsi que leur évaluation archivistique en fonction des critères définis. Ces interventions archivistiques représentent à la fois des particularités et s'avèrent complémentaires avec celles des autres spécialistes impliqués dans la gouvernance des jeux de données ouvertes de recherche. La description archivistique des jeux de données serait bénéfique pour retracer les activités ayant généré ceux-ci, notamment pour des besoins de vérification de la conformité aux critères de rigueur scientifique et comprendre l'influence de chaque processus de recherche sur la création, le traitement et l'exploitation de ces traces. La description archivistique pourrait s'avérer utile, par exemple, dans le cas des projets de recherche utilisant des devis méthodologiques mixtes séquentiels, où les données de la première phase de collecte servent à réaliser la seconde phase. Il en va de même pour les projets se réalisant en plusieurs phases, où les données de recherche sont liées entre elles par des relations de temporalité ou de finalités de recherche. L'absence des liens de temporalité, dont les métadonnées sur la création de chaque jeu de données, surtout pour les projets qui s'étalent sur plusieurs années (basés sur des méthodes longitudinales) guidant les usagers dans leur exploitation de ces données pour restituer l'activité de la recherche, ou encore la phase du projet dont il est question, pourrait rendre difficile l'usage

de ces jeux de données. La contextualisation soutient l'interprétation correcte des jeux de données, ce qui minimise les risques de leur mauvais usage.

Certains chercheurs, tels que Pasquetto et ses collaborateurs (2024), ont abordé les risques du mauvais usage des données de recherche, faute de leur contextualisation appropriée. Ils soulignent l'impact de ce mauvais usage sur la propagation des fausses informations, ce qui pourrait aller à l'encontre des ambitions recherchées par la science ouverte (Pasquetto et al., 2024). Cela va de pair avec les enjeux actuels de l'usage des données de recherche dans le domaine de l'apprentissage machine (Pasquetto et al., 2024; Paullada et al., 2021), dont la contextualisation s'avère nécessaire en explicitant les liens qui existent entre les jeux de données utilisés à des fins d'apprentissage machine, et ce, aux différentes étapes de leur création, traitement, diffusion et usage (Paullada et al., 2021). Si le plan de gestion des données fournit des détails au sujet du contexte de création de ces traces, il s'avère tout de même d'en extraire des métadonnées pour mieux contextualiser ces données. Cela ferait en sorte que les données soient mieux intelligibles et découvrables pour l'utilisateur, et ce, en minimisant les efforts déployés par ce dernier en vue d'accéder aux données qui l'intéressent.

Cela étant dit, l'intervention de l'archiviste universitaire serait essentielle dans une telle optique, notamment en agissant en amont, dès que les chercheurs ont des données complètes, afin de les guider dans la contextualisation de leurs traces et dans l'établissement des liens organiques qui existent entre elles, soutenant ainsi la transparence de la recherche scientifique. La contextualisation de ces jeux de données d'un point de vue archivistique peut se faire en utilisant plusieurs métadonnées archivistiques, comme celles de tout/partie (ex. *fait partie de*, ou *a pour partie*) pour illustrer les liens qui existent entre les jeux de données qui traitent de la même activité, notamment dans le cadre des projets utilisant des devis méthodologiques mixtes ou se réalisant en différentes phases. L'usage de cette métadonnée pourrait compléter celle portant sur le processus ou l'activité, ce qui favoriserait une meilleure traçabilité des activités de recherche. En outre, puisque les jeux de données de recherche sont souvent agrégés à d'autres objets documentaires (ex. jeux de données, pour en créer de nouveaux) dans différents contextes, l'intervention archivistique aiderait aussi à expliciter les différentes réutilisations possibles de ces traces, même par leurs producteurs, dans le cadre de différentes activités de recherche.

En clair, l'intégration de la perspective archivistique dans la contextualisation des jeux de données de recherche soutiendrait l'intelligibilité de ceux-ci comme preuves des activités scientifiques, en complétant ainsi la description bibliothéconomique qui se focalise sur la description du contenu et du contenant de ces traces. Les archivistes et les bibliothécaires universitaires auraient intérêt à fédérer leur expertise afin de mieux accompagner les producteurs dans la gestion de ces données, notamment leur description à la fois comme publications et comme traces. L'intervention archivistique ne se limiterait toutefois pas à la description de ces jeux de données, mais la dépasserait pour couvrir également leur évaluation pour en déterminer l'utilité pour les chercheurs et l'ensemble des usagers qui pourraient s'y intéresser.

Qui décide de la valeur que les jeux de données ouvertes de recherche possèdent? Les chercheurs, en tant que producteurs, se positionnent comme étant les acteurs les mieux placés pour définir les valeurs renfermées par ces données (Alaoui, 2024; Da Sylva, 2021; Makhoul-Shabou et Echernier, 2017). Sans surprise, ce sont les mieux placés pour en faire une telle évaluation, puisqu'ils connaissent bien les circonstances ayant conditionné

l'existence de ces traces. Cependant, les jeux de données ouvertes de recherche sont réutilisés non seulement par leurs producteurs eux-mêmes, mais aussi par d'autres usagers, pour les mêmes fins précédemment mentionnées. Les producteurs se basent sur plusieurs critères en vue de définir les délais de conservation de leurs jeux de données. D'une part, les recommandations des comités d'éthique de la recherche et des bureaux de recherche des universités, spécifiés notamment dans les formulaires de consentement et les certificats d'éthique délivrés aux chercheurs, pour des raisons en lien avec la protection de la vie privée des participants et la gouvernance de leurs données personnelles (Da Sylva, 2017; Da Sylva, 2021). D'autre part, des facteurs sont pris en considération pour la définition de ces délais de conservation, à savoir les besoins des producteurs eux-mêmes, que cela soit pour la conduite des activités de recherche ayant généré ces jeux de données, ou encore à des fins de vérifications ultérieures.

Ces interventions renvoient à la première étape de l'évaluation des données par les producteurs eux-mêmes. Une fois les données de la recherche ne servent plus à atteindre les besoins de ces acteurs, elles subissent un deuxième processus d'évaluation archivistique. Dans cette optique, plusieurs critères sont pris en considération par les archivistes, dont l'intention de préservation, la pertinence par rapport à la mission de l'organisation conservatrice de ces données (ici l'université), le potentiel de la réutilisation, les contraintes légales et réglementaires, la valeur archivistique (scientifique ou historique), la question des coûts, la qualité des données (l'unicité) et leur complétude (Dorey et al., 2022). Cette méthodologie d'évaluation archivistique des données de recherche s'avère relativement coûteuse, puisque ces dernières subissent deux étapes d'évaluation. Qui plus est, les critères archivistiques se voient limités à ceux à caractère scientifique et historique, ce qui correspond à la valeur secondaire des jeux de données comme documents. Cette dernière se définit *après* la valeur primaire, laquelle est déterminée par les producteurs eux-mêmes pour leurs besoins de conduite du projet de recherche en lien avec les données. Cependant, ces traces sont utilisées d'une manière simultanée par le même chercheur pour diverses fins, ce qui laisse entendre qu'elles peuvent renfermer plus qu'une seule valeur archivistique, dépassant celles à caractère scientifique et historique. Considérant le triple pouvoir des jeux de données de recherche comme documents, les chercheurs peuvent les utiliser à des fins institutionnelles, organisationnelles et communautaires.

En archivistique, la théorie des valeurs des documents considère que les documents ont soit une valeur primaire ou secondaire (Schellenberg, 1956). La valeur primaire des documents correspond à l'usage de ces derniers à des fins administratives, financières, légales et fiscales, tandis que la valeur secondaire renvoie à leur exploitation pour des fins historiques, scientifiques, patrimoniales et d'information générale (Schellenberg, 1956). Or, une telle théorie a été élaborée dans un contexte bureaucratique, où les documents d'archives servent d'abord et avant tout des finalités administratives et légales. Cette lecture fait de la valeur scientifique des documents une valeur secondaire, ce qui la rend moins adaptée au contexte de la recherche scientifique. En fait, dans un contexte universitaire, les missions qui sont en jeu sont l'enseignement et la recherche, ce qui laisse entendre que les documents produits ou reçus par les chercheurs possèdent d'emblée une valeur primaire de nature scientifique. Ainsi, force est d'admettre que cette lecture des valeurs archivistiques, telle que théorisée par Schellenberg (1956), est moins pertinente pour la définition des valeurs que les jeux de données ouvertes de recherche peuvent avoir. Une lecture fonctionnelle des valeurs archivistiques nous aiderait à définir les différentes valeurs que ces traces documentaires possèdent.

Dans sa critique des valeurs archivistiques, Cardin (1994) propose une lecture fonctionnelle de celles-ci, sous trois angles, à savoir l'angle probatoire, opérationnel et symbolique. D'après l'auteure, la valeur probatoire des documents d'archives peut être de nature administrative, légale ou scientifique. La valeur de preuve administrative, correspondant à celle de Schellenberg (1956), renvoient aux rôles des documents comme preuves de l'activité administrative. La valeur de preuve légale illustre la capacité des documents à incarner la dimension juridique qui atteste la véracité des faits par une autorité compétente (Cardin, 2021). Pour sa part, la valeur de preuve scientifique, qui nous intéresse d'ailleurs, renvoie à la capacité des documents à fournir une démonstration scientifique objective, conforme aux critères de rigueur scientifique reconnus au sein des communautés scientifiques (Cardin, 2021). Cette valeur de preuve scientifique s'aligne bien avec les finalités d'usage des jeux de données ouvertes de recherche, puisqu'elle est celle qui prime lorsque les producteurs et les usagers de ces données exploitent celles-ci à des fins de recherche scientifique, notamment en vérifiant la véracité des faits et leur conformité aux critères de rigueur collectivement reconnus par l'institution scientifique canadienne. Parmi ces derniers, pensons à la qualité de la démonstration scientifique et sa conformité avec les normes établies dans chacune des disciplines (sciences humaines et sociales, sciences exactes et sciences de la santé). Cela correspond à la finalité d'exploitation des données par les pairs, pour contrer les risques de falsification et de fabrication des données, lesquels s'avèrent des enjeux auxquels se heurtent les pratiques de communication savante (Nazarovets, 2024).

Quant à la valeur archivistique opérationnelle, elle renvoie à la capacité des documents à combler des besoins courants, factuels et de vérification des données opérationnelles (Cardin, 1994). Dans l'optique de la recherche scientifique, les jeux de données ouvertes de recherche possèdent une valeur opérationnelle qui permet aux producteurs et aux usagers de ces traces de se renseigner sur des aspects factuels tels que les noms des chercheurs, des organismes subventionnaires, les universités et les centres de recherche d'attache, etc. Cette valeur soutient les producteurs eux-mêmes dans la recherche des informations factuelles sur leurs propres projets de recherche, comme elle pourrait aider les usagers à se faire une idée sur les types de projets de recherche subventionnés, leur discipline, ainsi que d'avoir un aperçu sur les universités dont les équipes de recherche se voient attribuer le plus grand nombre de subventions. Cela pourrait aider à positionner les universités canadiennes sur le plan du rendement scientifique et du financement de la recherche, de même que leur visibilité en tant qu'organisations en quête de la génération de la valeur ajoutée en attirant des étudiants de tous les coins du monde et recherchant des opportunités de financement de différents projets de recherche par des organismes reconnus. Cette valeur archivistique est souvent celle à laquelle les bureaux de recherche et les agents administratifs s'intéressent, notamment dans leur contexte quotidien d'administration des programmes et des subventions de recherche. Cette valeur opérationnelle renvoie aussi à l'utilité de ces jeux de données lorsqu'ils permettent de vérifier des faits relatifs à leur contenu, indépendamment de leur valeur probatoire. Par exemple, il pourrait être question, pour un chercheur universitaire, du simple fait de se renseigner sur le nombre d'étudiants ayant indiqué qu'ils sont d'accord sur le fait que la formation à distance est plus bénéfique que celle en présentiel, et ce, dans le cadre d'un sondage mené par un chercheur en sciences de l'éducation s'intéressant aux bénéfices de la formation à distance en milieu universitaire.

Enfin, la valeur symbolique des documents renvoie à la capacité de ces derniers à agir à titre de témoins de l'évolution du discours identitaire (ex. traditions, valeurs, principes, idéologies) de la société (Cardin, 2021). Dans le contexte de la recherche, les chercheurs universitaires sont souvent affiliés à des groupes et équipes de recherche par discipline, ou encore réunissant plusieurs champs disciplinaires pour encourager la recherche interdisciplinaire. Ces groupes partageant des orientations communes, des valeurs et des principes de rigueur scientifique et de liberté académique. Pourvu que ces valeurs soient reproductibles et transmissibles d'une communauté de pairs à une autre, plusieurs documents sont mobilisés (Alaoui, 2024). Les jeux de données ouvertes de recherche, en tant que véhicules de ces traits identitaires, sont porteurs d'une valeur symbolique. Ils retracent l'évolution du discours scientifique, notamment en militant pour une meilleure transparence et accessibilité des traces de la recherche. Ils témoignent en outre de l'évolution des normes d'évaluation de la recherche fondamentale et appliquée, puisque les normes reconnues comme étant valides au sein de ces communautés changent au fil du temps. Aussi, les sujets auxquels les chercheurs (producteurs et usagers) s'intéressaient auparavant ne sont pas nécessairement les mêmes que ceux d'aujourd'hui, ce qui confirme, encore une fois, le rôle des jeux de données dans la description des thématiques de recherche intéressant la communauté académique canadienne au fil des années.

Cette lecture fonctionnelle des valeurs archivistiques aide à établir une certaine distanciation par rapport à la lecture traditionnelle des valeurs et leur découpage selon lequel la valeur secondaire se définit lors d'une réévaluation des documents, après la détermination de leur valeur primaire. Dans une telle optique fonctionnelle, il faut considérer les valeurs archivistiques plurielles de ces jeux de données ouvertes. Celles-ci sont concurrentes (Cardin, 1994), car elles sont présentes d'une manière simultanée dans un même jeu de données ouvertes de recherche, peu importe son stade de vie. Cette interprétation des valeurs archivistiques d'un point de vue fonctionnel pourrait aider à anticiper, dès la création des données de recherche, leurs finalités d'usage à la fois par les producteurs et les usagers, en définissant les usages primaires et secondaires de ces données. Cela pourrait minimiser les efforts de la double évaluation des jeux de données d'un point de vue archivistique et en élargir les sphères d'usage.

Considérant la diversité des finalités d'exploitation de ces données ouvertes de recherche et le caractère pluriel de leur valeur archivistique, il importe que celle-ci s'opère dans une perspective collaborative, mettant en jeu les producteurs des données, les experts en gestion de l'information et les représentants des usagers comme parties prenantes clé. D'un point de vue archivistique et à la lumière des considérations actuelles en lien avec l'ouverture et l'inclusion des usagers dans l'adaptation de l'offre de l'information documentaire en ligne, il importe que ces derniers soient impliqués dans les différentes étapes de la chaîne archivistique (Cardin et Desilets, 2018). La définition des valeurs archivistiques plurielles de ces jeux de données ouvertes de recherche devrait se faire dans une perspective collaborative, impliquant des démarches de dialogue et de médiation culturelle entre les producteurs des données et leurs usagers, de concert avec un ensemble de spécialistes (Alaoui, 2024). Cela aiderait l'ensemble des parties concernées à se mettre d'accord sur les valeurs archivistiques des données et, par conséquent, établir leurs délais de conservation et leur sort final, soit la conservation permanente ou l'élimination.

Le dialogue entre les parties prenantes pour une meilleure définition collaborative de la valeur archivistique plurielle des jeux de données ouvertes de recherche

Reconnaissant l'importance de la collaboration dans le développement d'un cadre co-construit de la gouvernance des jeux de données ouvertes de recherche et en réponse aux dispositions de la *Politique des trois organismes subventionnaires sur la gestion des données de recherche*, les universités canadiennes se sont dotées de leurs propres stratégies en gestion des données de recherche. Ces stratégies sont développées grâce à la collaboration entre les divers membres des comités de gestion des données ouvertes de recherche. Ces comités consultatifs sont souvent composés de représentants des professeurs, des bureaux de recherche (ex. conseillers en développement et en valorisation de la recherche), des experts en TI, des analystes des données, des conseillers en conformité légale, ainsi que des bibliothécaires universitaires spécialisés en gestion des données ouvertes de recherche. À ces acteurs s'ajoutent des représentants des étudiants aux cycles supérieurs (maîtrise et doctorat), ainsi que des professionnels de recherche et des stagiaires postdoctoraux représentant les trois grands secteurs de la recherche (c'est-à-dire les sciences humaines et sociales, les sciences et génie, et les sciences de la santé). Ces parties prenantes sont les plus impliquées dans la gestion des jeux de données ouvertes de recherche, puisque les pratiques de la communication savante les interpellent en premier chef. Cependant, qu'en est-il des archivistes universitaires? Ces derniers sont, le plus souvent, absents de cette cartographie des acteurs siégeant sur ces comités consultatifs, malgré leur apport indéniable à la gouvernance de ces jeux de données de recherche. S'ils sont présents, les rôles de ces spécialistes se limitent souvent aux questions d'archivage à long terme¹⁵, notamment en définissant les critères de préservation pérenne et les formats technologiques favorisant la persistance des jeux de données de recherche. Cependant, les archivistes universitaires auraient intérêt à revendiquer une plus grande visibilité de leur rôle dans la gestion des données ouvertes de recherche, notamment en prouvant la manière dont ils peuvent apporter une valeur ajoutée à cette gestion, en renforçant ainsi la conformité aux principes FAIR.

En s'alliant aux groupes de travail sur la gestion des données de recherche, ou encore aux comités institutionnels en lien avec une telle activité, les archivistes universitaires auraient la possibilité d'accompagner les producteurs à toutes les étapes de la gestion du cycle de vie des données qu'ils produisent dans le cadre de leurs projets. En s'inspirant des démarches entreprises par les bibliothécaires, ils pourraient enrichir les modèles de plans de gestion des données mis à la disposition des producteurs par des aspects archivistiques liés à la contextualisation de ces données en utilisant des métadonnées pour enrichir les liens à établir entre celles-ci comme produits de l'activité de recherche. De surcroît, ils élaboreraient des guides et des manuels pour la définition des valeurs archivistiques plurielles de ces données. Ces documents comporteraient des explications de la démarche

¹⁵ Voir, à ce sujet, le cas des comités de l'Université de Montréal (https://recherche.umontreal.ca/fileadmin/recherche/documents/pdf/Strategie_institutionnelle_GDR_UdeM.pdf), de l'Université du Québec à Montréal (<https://gdr.uqam.ca/strategie/gouvernance-coordination-et-offre-de-services/>) et de l'Université de l'Ontario français (https://uontario.ca/fileadmin/user_upload/Strategie-UOF_Gestion_donnees_de_recherche_2024-02-13.pdf)

pour déterminer les valeurs archivistiques d'un point de vue fonctionnel, notamment en répondant à un ensemble de questions (voir tableau 1).

Valeur archivistique	Questions à poser
<i>Valeur probatoire</i>	<ul style="list-style-type: none"> • Les jeux de données reflètent-ils la démonstration de la méthodologie de la recherche d'une manière neutre? Répondent-ils aux critères de rigueur scientifique? • Les liens entre les jeux de données permettent-ils de retracer l'activité de recherche les ayant générés? • Est-il possible de retracer les différentes phases du projet en se fiant aux jeux de données et à leurs liens organiques?
<i>Valeur opérationnelle</i>	<ul style="list-style-type: none"> • Les jeux de données permettent-ils de renseigner sur des aspects factuels? • Ces jeux de données sont-ils utiles à des vérifications administratives?
<i>Valeur symbolique</i>	<ul style="list-style-type: none"> • Les jeux de données sont-ils porteurs de significations identitaires et symboliques, par exemple l'évolution des valeurs, des discours et des pratiques de recherche scientifique, des relations entre les universités et la société, et des rapports entre les acteurs sociaux?

Tableau 1 Exemples de questions à se poser pour déterminer les valeurs archivistiques des jeux de données de recherche

Comment un tel processus pourrait-il s'opérer ? Nous suggérons une démarche proactive dans l'évaluation de ces données. Les archivistes universitaires pourraient mener des activités de consultation auprès de ces usagers pour cerner leurs attentes en termes d'accès aux données. Quelle est la portée des données qui les intéresse le plus ? Quelles en sont les caractéristiques et les critères de qualité notamment archivistique? Pour ce faire, il serait important de nouer des relations de dialogue avec les représentants des usagers, par exemple les porte-paroles des associations étudiantes, des professionnels de recherche, des stagiaires postdoctoraux, ou encore des agents administratifs, en vue de cerner leurs besoins. Après la compilation des besoins, les archivistes universitaires pourraient établir une priorisation des jeux de données en fonction de leur portée. Des suggestions seraient faites par ces experts auprès du groupe de travail sur la gestion des données de recherche, et ce, pour trouver des solutions qui répondent aux préoccupations de toutes les parties. Il est important de garder en tête que les jeux de données qui sont libérés sur des portails publics sont essentiellement ceux qui découlent des projets de recherche financés par les organismes subventionnaires de la recherche du Canada. Or, sur le terrain universitaire, ces traces sont également créées par des chercheurs qui ne sont pas titulaires de tels financements, ce qui commande une mobilisation de ces acteurs pour les inciter à déposer leurs jeux de données en ligne. Ainsi, les archivistes universitaires devraient participer, de concert avec les autres membres du comité consultatif, à la définition des stratégies motivant ces chercheurs pour déposer leurs jeux de données de recherche en ligne. Il pourrait être question de la mise en place d'un système de récompenses pour aider les dépositaires des données à gagner en visibilité, en pensant, par exemple, à la publication de leurs noms et au partage des détails de leurs projets de recherche sur le site de la

bibliothèque de l'université. Somme toute, le dialogue entre l'ensemble des parties prenantes est nécessaire pour faire de la libération des données ouvertes de recherche une démarche renforçant non seulement le libre accès à la recherche. Il favorise en outre la prise en compte des contraintes de chaque acteur, notamment en lien avec la protection de la vie privée, les compétences, le respect du droit d'auteur, de même que la question des ressources nécessaires à la diffusion proactive de ces données en ligne.

Limites de la recherche

L'objectif de cet article théorique était d'aboutir à une nouvelle conceptualisation du rôle de l'archiviste universitaire dans la gouvernance des jeux de données ouvertes de recherche. Pour ce faire, nous nous sommes basée sur une critique de sources secondaires, comme les corpus scientifiques relevant de la littérature en sciences de l'information, couplée à une consultation des sites web de certaines institutions canadiennes. Ainsi, les conclusions de notre réflexion théorique devraient faire l'objet d'une expérimentation sur le terrain universitaire pour valider le rôle des archivistes dans la gouvernance des données. Cela aiderait en outre à identifier les conditions de l'intégration de la perspective archivistique dans cette gouvernance, notamment en cernant les sources des conflits avec les autres parties prenantes et en explorant les pistes de solutions pour élargir la sphère d'intervention des archivistes dans une telle pratique collaborative.

Conclusion

Dans cet article, nous avons abordé la gestion collaborative des jeux de données de recherche en vue de leur libération sur des portails publics des universités canadiennes, et ce, pour soutenir le mouvement de la science ouverte. Plus spécifiquement, nous nous sommes penchée sur la perspective archivistique de la gestion de ces traces documentaires granulaires, laquelle est peu explorée dans la littérature scientifique. D'une part, après avoir exposé les caractéristiques des jeux de données de recherche en tant que preuves des activités de recherche universitaire, nous avons positionné les acteurs universitaires dans le large contexte de la gestion de ces preuves documentaires. Le rôle des archivistes y a été également mis de l'avant, notamment en regard de la contextualisation de ces jeux de données pour être en mesure de retracer les activités de recherche ayant conditionné la création de ces derniers. Il est aussi question de l'évaluation de ces jeux de données en tant que documents, et ce, en posant un nouveau regard sur les valeurs archivistiques. Celles-ci ont été étudiées sous l'angle fonctionnel, notamment en mettant de l'avant leurs aspects probatoire, opérationnel et symbolique. L'idée était de dépasser la lecture traditionnelle des valeurs archivistiques pour mettre de l'avant les valeurs plurielles des jeux de données de recherche, lesquelles correspondent aux finalités simultanées d'usage de ces derniers en fonction des schèmes de pratique des usagers. Enfin, nous avons souligné l'importance de la collaboration entre les producteurs de données, les experts impliqués dans leur gestion, ainsi que les usagers, notamment par le biais de comités institutionnels comportant des représentants de ces parties prenantes. Cela stimulerait le dialogue entre l'ensemble des parties, de telle sorte que leurs besoins et attentes soient pris en considération pour établir le juste équilibre entre les ambitions de la science ouverte et les contraintes éthiques et légales, de même que les compétences et les ressources nécessaires à la libération des jeux de données ouvertes de recherche. Il reste à voir comment les archivistes pourraient

s'imposer dans un tel contexte, notamment en faisant valoir leur apport à l'usage de ces jeux de données comme traces authentiques, fiables, intègres et exploitables, en essayant de sortir de leur ombre pour tourner leur rôle davantage vers l'externe.

Acknowledgments

Rien à signaler.

Competing interests

Aucun conflit d'intérêt.

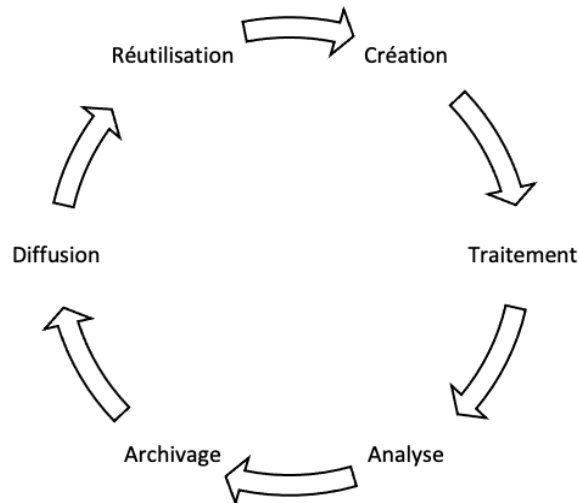
Tableaux et figures

Tableau 1 Exemples de questions à se poser pour déterminer les valeurs archivistiques des jeux de données de recherche..... 113

Valeur archivistique	Questions à poser
<i>Valeur probatoire</i>	<ul style="list-style-type: none"> • Les jeux de données reflètent-ils la démonstration de la méthodologie de la recherche d'une manière neutre? Répondent-ils aux critères de rigueur scientifique? • Les liens entre les jeux de données permettent-ils de retracer l'activité de recherche les ayant générés? • Est-il possible de retracer les différentes phases du projet en se fiant aux jeux de données et à leurs liens organiques?
<i>Valeur opérationnelle</i>	<ul style="list-style-type: none"> • Les jeux de données permettent-ils de renseigner sur des aspects factuels? • Ces jeux de données sont-ils utiles à des vérifications administratives?
<i>Valeur symbolique</i>	<ul style="list-style-type: none"> • Les jeux de données sont-ils porteurs de significations identitaires et symboliques, par exemple l'évolution des valeurs, des discours et des pratiques de recherche scientifique, des relations entre les universités et la société, et des rapports entre les acteurs sociaux?

Figure 1 Le cycle de vie des données de recherche (Inspiré de Higgins, 2012)..... 105

Figure 2 Liste des métadonnées à associer aux jeux de données de recherche en vue de leur publication sur *Borealis* (Borealis Université Laval, s.d.) 107



borealis

- Identifiant pérenne ?
- Date de publication ?
- Titre ?
- Autorat ?
- Personne-ressource ?
- Description ?
- Sujet ?
- Mot-clé ?
- Classification par sujets ?

- Classification par sujets ?
- Publication connexe ?
- Renseignements sur la subvention ?
- Dépôt par ?
- Date de dépôt ?
- Date de la collecte ?

Références

Alaoui, S. (2024). Modélisation de la gestion de l'information dans une perspective ouverte : le cas d'une université québécoise en transformation numérique [Thèse de doctorat, Université Laval]. Corpus UL. <https://corpus.ulaval.ca/entities/publication/74c064a8-1faf-49f6-bf6a-388f72a54449>

Awada, L., Phillips, P. W. B., et Bogdan, A. M. (2022). Governance and stewardship for research data and information sharing : Issues and prospective solutions in the transdisciplinary plant

- phenotyping and imaging research center network. *Plants, People, Planet*, 4(1), 84-95.
<https://doi.org/10.1002/ppp3.10238>
- Barsky, E., Hill, E., Zaraiskaya, T., Wang, M. et Costanzo, L. (2023). Historique et paysage canadien de la gestion des données de recherche.
<https://ecampusontario.pressbooks.pub/gdrCanada/chapter/historique-et-paysage-canadien-de-la-gdr/>
- Borealis Data. (s.d.). Borealis : Le dépôt Dataverse canadien.
<https://dataverse.scholarsportal.info/fr/>
- Borealis (s.d.). Borealis : le Dataverse de l'Université Laval.
<https://borealisdata.ca/dataverse/laval>
- Borgerud, C., et Borglund, E. (2020). Open research data, an archival challenge?. *Archival Science*, 20(3), 279-302. <https://link.springer.com/article/10.1007/s10502-020-09330-3>
- Cardin, M. (1994). Information, preuve et témoignage ou le triple pouvoir des archives. Dans *Les valeurs archivistiques. Théorie et pratique. Actes du colloque organisé - 388 - conjointement par la Division des archives et les Programmes d'archivistique de l'Université Laval* (pp. 7-24). Université Laval, Québec, Canada.
- Cardin, M. (2013-2014). Penser l'exploitation des archives en tant que système complexe. *Archives*, 45(1), 135- 146. https://www.archivistes.qc.ca/revuearchives/vol45_1/45_1_cardin.pdf
- Cardin, M. (2021). Cours 2 : Les valeurs archivistiques [notes de cours GAD 2108- Diffusion et valorisation des archives]. Département des sciences historiques, Université Laval.
<https://sitescours.monportail.ulaval.ca/>
- Cardin, M. et Desilets, C. (2018). La coagrégation des contenus en publicité : un premier laboratoire pour l'archivistique ouverte. Dans M. Cardin et A. Klein (dir.). *Consommer l'information : de la gestion à la médiation documentaire* (7-26). Presses de l'Université Laval.
- Castets-Renard, C., et Gandon, N. (2016). Open data des données de la recherche publique : Entre réformes législatives et retour d'expérience sur un guide pratique à destination des chercheurs. *Legicom*, 56(1), 67. <https://doi.org/10.3917/legi.056.0067>
- Couture, C., et Roy, J. (2006). La norme ISO 15489: principes et application. *Archives*, 38(2), 2006-2007.
- Da Sylva, L. (2021). Le rôle des acteurs institutionnels dans la gestion des données de recherche. Présentation dans le cadre du Colloque sur la gestion des données de recherche en sciences humaines, 7 et 8 juin 2021, Université Laval.
- Da Sylva, L., Maurel, D., Bruyère, M., Saint-Germain, M., et Gareau, G. (2019). Écosystème de la gestion de données de recherche et professionnels de l'information : présentation des enjeux, de la méthodologie et des solutions préconisées d'une enquête canadienne. *Études de communication* (52), 51-70. <https://journals.openedition.org/edc/8615>
- Dillaerts, H. (2020). Usages et pratiques en lien avec les données de recherche. Une enquête menée auprès des chercheurs de l'université Paul-Valéry Montpellier 3. <https://halshs.archives-ouvertes.fr/halshs-02902710/document>
- Dorey, J., Hurllet, G. et Knazook, B. (2022). Guide d'évaluation pour la préservation des données de recherche. Zenodo. <https://doi.org/10.5281/zenodo.6283886>
- Duranti, L., et Thibodeau, K. (2006). The Concept of Record in Interactive, Experiential and Dynamic Environments : The View of InterPARES*. *Archival Science*, 6(1), 13- 68.
<https://doi.org/10.1007/s10502-006-9021-7>

Gouvernement du Canada (2014). Plan d'action du Canada pour un gouvernement ouvert 2014-2016. <https://ouvert.canada.ca/fr/contenu/plan-daction-du-canada-gouvernement-ouvert-2014-2016>

Gouvernement du Canada (2021). Politique des trois organismes sur la gestion des données de recherche. <https://science.gc.ca/site/science/fr/financementinterorganismes-recherche/politiques-lignes-directrices/gestion-donneesrecherche/politique-trois-organismes-gestion-donnees-recherche>

Gouvernement du Canada (s.d.). Plan d'action national pour un gouvernement ouvert. <https://www.canada.ca/fr/gouvernement/systeme/rapports-pangouvernementaux-depenses-activites/confiance-transparence/a-propos-gouvernement-ouvert/plan-action-national-gouvernement-ouvert.html>

Gouvernement du Canada (s.d.). Plan d'action pour la science ouverte : 2021-2026. <https://www.canada.ca/fr/environnement-changement-climatique/services/sciences-technologies/plan-action-science-ouverte.html>

Gouvernement du Canada. (s.d.). Politique des trois organismes sur le libre accès aux publications. <https://science.gc.ca/site/science/fr/financement-interorganismesrecherche/politiques-lignes-directrices/libre-acces/politique-trois-organismes-libreacces-aux-publications>

International research on permanent authentic records in electronic systems (InterPARES2) (s.d.). Au glossaire du projet InterPARES2.

http://www.interpares.org/ip2/display_file.cfm?doc=ip2_livre_glossaire.pdf

Higgins, S. (2012). The lifecycle of data management. Dans G. Pryor (Ed.), *Managing Research Data* (pp. 17–46). Facet Publishing.

Jacquemin, B., Schöpfel, J., et Fabre, R. (2019a). Libre accès et données de recherche. De l'utopie à l'idéal réaliste. *Études de communication*, 52, 11-26. <https://doi.org/10.4000/edc.8468>

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Lapointe, J.-M., et Castonguay, R. (2020). Les bibliothécaires universitaires, pédagogues de l'écosystème socionumérique de l'information : Enjeux et opportunités d'une mutation professionnelle en cours. *I2D - Information, données et documents*, 1(1), 46.

<https://doi.org/10.3917/i2d.201.0046>

Makhlouf-Shabou, B., et Echernier, L. (2017). La gouvernance du cycle de vie des données de la recherche en Suisse : Transfert et pérennisation. *La Gazette des archives*, 246(2), 89-100.

<https://doi.org/10.3406/gazar.2017.5497>

Mintzberg, H. (1982). *Structure et dynamique des organisations*. Éditions d'organisation.

Nazarovets, S. (2024). Dealing with Research Paper Mills, Tortured Phrases, and Data Fabrication and Falsification in Scientific Papers. In *Scientific Publishing Ecosystem: An Author-Editor-Reviewer Axis* (pp. 233-254). Singapore: Springer Nature Singapore.

https://link.springer.com/chapter/10.1007/978-981-97-4060-4_14

OCDE. (2007). Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics. https://www.oecd-ilibrary.org/science-andtechnology/oecd-principles-and-guidelines-for-access-to-research-data-frompublic-funding_9789264034020-enfr;jsessionid=EocBIhdhquS8p14WQFjztn1eWdfu1MI dj2UEof6cu.ip-10-240-5-107

Open Government Partership (s.d.). À propos. <https://www.opengovpartnership.org/fr/>

- Pasquetto, I. V., Cullen, Z., Thomer, A., & Wofford, M. (2024). What is research data “misuse”? And how can it be prevented or mitigated?. *Journal of the Association for Information Science and Technology*, 75(12), 1413-1429.
- Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis) contents: A survey of dataset development and use in machine learning research. *Patterns*, 2(11), 100336.
- Portage Network (2020). Brief Guide - Data Management Plan. Zenodo. <https://doi.org/10.5281/zenodo.4495482>
- Quarati, A., et Raffaghello, J. E. (2022). Do researchers use open research data? Exploring the relationships between usage trends and metadata quality across scientific disciplines from the Figshare case. *Journal of Information Science*, 48(4), 423-448. <https://doi.org/10.1177/0165551520961048>
- Reymonet, N. (2017). Améliorer l'exposition des données de la recherche : La publication de data papers. https://archivesic.ccsd.cnrs.fr/sic_01427978/document
- Sargiotis, D. (2024). Overview and Importance of Data Governance. Dans *Data Governance: A Guide* (pp. 1-85). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-67268-2_1
- Schellenberg, T. R. (1956). *Modern archives*. University of Chicago Press.
- Senécal, S. (2007-2008). Documents et connaissances. *Archives*, 39(2), 89-107. https://www.archivistes.qc.ca/revuearchives/vol39_2/39_2_senecal.pdf
- Shepherd, E., et Yeo, G. (2013). *Managing Records : A handbook of principles and practice* (1re éd.). Cambridge University Press. <https://doi.org/10.29085/9781856049788>
- Toulouse, J. M. (2007). Rapport de recherche sur la gouvernance des institutions universitaires. HEC Montréal. https://spuqat.uqat.ca/documents/gouvernance/Rapport_de_recherche_sur_la_gouvernance_des_institutions_universitaires.pdf
- UNESCO. (s.d.). Open Science: Making science more accessible, inclusive and equitable for the benefit of all. <https://www.unesco.org/en/open-science>
- Université de l'Ontario français (2023). Stratégie institutionnelle de gestion des données de recherche. https://uontario.ca/fileadmin/user_upload/Strategie-UOF_Gestion_donnees_de_recherche_2024-02-13.pdf
- Université de Montréal (2024). Stratégie institutionnelle pour la gestion des données de recherche. https://recherche.umontreal.ca/fileadmin/recherche/documents/pdf/Strategie_institutionnelle_GDR_UdeM.pdf
- Université du Québec à Montréal (s.d.). Stratégie institutionnelle pour la gestion des données de recherche. <https://gdr.uqam.ca/strategie/preambule/>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>